

ВЫБОР ПРЕДСТАВЛЕНИЙ В РЕЛЯЦИОННЫХ БАЗАХ ДАННЫХ И ХРАНИЛИЩАХ ДАННЫХ

VIEWS SELECTION IN RELATIONAL DATABASES AND DATA WAREHOUSES

A. Conde

Summary. Using views in databases of various types can be useful in a wide variety of situations. This article analyses the most well known heuristic algorithms proposed in the literature to solve the problem of views selection. The methods of views selection that have been proposed in the literature in order to optimize databases are discussed.

Keywords: view, view selection, heuristic algorithms, DBMS, materialized views, view selection methods.

Конде Абдул Карим

Аспирант, Московский государственный технологический университет «СТАНКИН», г. Москва
karimconde112@gmail.com

Аннотация. Использование представлений в базах данных различных типов может оказаться полезным в самых разнообразных ситуациях. В данной статье анализируются наиболее известные эвристические алгоритмы, предложенные в литературе для решения задачи выбора представления. Обсуждаются методы выбора представления, которые были предложены в литературе с целью оптимизации баз данных.

Ключевые слова: представление, выбор представления, эвристические алгоритмы, СУБД, материализованные представления, методы выбора представления.

Представление — это виртуальная таблица, содержание которой определяется запросом. Как и таблица, представление состоит из набора именованных столбцов и строк данных. Если представление не проиндексировано, оно не существует как сохраненный набор значений данных в базе данных.

Представление действует как фильтр на базовые таблицы, на которые ссылается представление. Запрос, определяющий представление, может быть из одной или нескольких таблиц или из других представлений в текущей или других базах данных. Распределенные запросы также можно использовать для определения представлений, которые используют данные из нескольких разнородных источников.

Представления обычно используются для фокусировки, упрощения и настройки восприятия базы данных каждым пользователем. Представления могут использоваться в качестве механизмов безопасности, позволяя пользователям получать доступ к данным через представление, не предоставляя им прав на прямой доступ к базовым таблицам, лежащим в основе представления. Представления могут использоваться для обеспечения обратного совместимого интерфейса для эмуляции таблицы, которая существовала ранее, но схема которой изменилась.

Выбор представления — это задача, которая состоит из выбора набора представлений, которые должны быть материализованы, чтобы повысить производительность запроса. Мы используем термин «выбор представления» как взаимозаменяемый с термином «выбор материализованного представления».

Материализованное представление сохраняет данные, возвращаемые в результате запроса на определение представления, и автоматически обновляется при изменении данных в базовых таблицах. Оно повышает производительность сложных запросов (обычно запросов с объединениями и агрегатами), предлагая при этом простые операции обслуживания. Благодаря возможности автоматического подбора плана выполнения материализованное представление не обязательно должно быть упомянуто в запросе, чтобы оптимизатор рассмотрел его для замены. Эта возможность позволяет инженерам по обработке данных использовать материализованные представления в качестве механизма для улучшения времени отклика запросов без необходимости изменения запросов.

Наиболее известные эвристические алгоритмы, предложенные в литературе для решения задачи выбора представления перечислены ниже:

1. Детерминированные алгоритмы: алгоритмы этого класса обычно строят решение детерминированным образом путем исчерпывающего поиска или применяя некоторые эвристики, такие как жадный алгоритм, чтобы избежать необходимости обхода пространства решений в исчерпывающем поиске. Однако жадный поиск имеет известные недостатки, т.е. вместо глобально оптимального решения могут быть оставлены субоптимальные решения, поскольку начальные решения сильно влияют на решение. Очень трудно найти оптимальное решение задач, которые относятся к классу NP-полных задач из-за того, что пространство решений растет экспоненциально по мере увеличения размера задачи. Например, в контексте проблемы

выбора представления, количество возможных представлений (комбинаций представлений) растет экспоненциально с количеством запросов в рабочей нагрузке, количеством столбцов, предикатов присоединения, пунктов группировки и таблиц, на которые ссылается каждый запрос, а также с количеством узлов компьютера, если проблема изучается в распределенном сценарии.

2. Рандомизированные алгоритмы: наиболее часто используемыми рандомизированными алгоритмами в контексте выбора представления являются алгоритмы имитационного отжига и генетические алгоритмы.

— Имитационные алгоритмы отжига: такие алгоритмы мотивированы аналогией с отжигом в твердых телах. Они основаны на технике итеративного улучшения, которая применяется к одной точке, представляющей собой решение в пространстве поиска, и постоянно пытается перебрать ее соседей, чтобы найти лучшую точку (лучшее решение). Для того чтобы устранить зависимость от начальной точки поиска, алгоритмы имитационного отжига используют вероятность принятия решения о том, следует ли переходить к соседней точке. Действительно, можно перейти в соседнюю точку (соседнее решение) путем случайной ходьбы, которая может быть дальше от оптимума, чем предыдущая, в ожидании, что ее соседи будут представлять лучшее решение. Вероятность принятия рассчитывается в соответствии с графиком охлаждения. Алгоритмы завершаются, как только не существует подходящих ходов или теряется вся энергия в системе.

— Генетические алгоритмы: эти алгоритмы генерируют решения, используя методы, вдохновленные процессом естественной эволюции, такие как отбор, мутация и кроссинговер. Поиск стратегии для этих алгоритмов очень похож на биологическую эволюцию. Генетические алгоритмы используют рандомизированную стратегию поиска; они начинают со случайной начальной популяции, содержащей особи, которые представляют возможные решения, и генерируют новые популяции путем случайного скрещивания и мутации. Найденная особь является решением. Алгоритмы завершаются, как только не происходит дальнейшего улучшения в течение определенного периода. В отличие от алгоритмов имитационного отжига, генетические алгоритмы используют многонаправленный поиск поддерживая пул точек кандидатов (решений-кандидатов) в пространстве поиска. Между точками кандидатами происходит обмен информацией для направления поиска, при котором хорошие кандидаты выживают, а плохие погибают. Этот многонаправленный эволюционный подход позволяет генетическому ал-

горитму эффективно искать в пространстве и находить точку, близкую к глобальному оптимуму.

Рандомизированные алгоритмы основаны на статистических концепциях, согласно которым пространство поиска может быть исследовано случайным образом до достижения точки, близкой к глобальному оптимуму. Это может быть применено для очень больших пространств поиска. Кроме того, они могут найти разумное решение за относительно короткий промежуток времени, обменивая время выполнения на качество. Однако нет никакой гарантии производительности, поскольку вероятностное поведение генетических алгоритмов не гарантирует нахождения глобального оптимума.

3. Гибридные алгоритмы: гибридные алгоритмы сочетают в своем поиске стратегии чисто детерминированных алгоритмов и чисто рандомизированных алгоритмов, чтобы обеспечить лучшую производительность с точки зрения качества решения. Решения, полученные детерминированными алгоритмами, используются в качестве начальной конгруэнции для алгоритмов имитационного отжига или в качестве начальной популяции для генетических алгоритмов. Комбинация возможностей рандомизированных и детерминированных алгоритмов может обеспечить лучшее качество решения, чем рандомизированные или детерминированные алгоритмы, используемые по отдельности. Однако гибридные алгоритмы более трудоемки, так как в процессе поиска приходится тратить значительное количество времени. Поэтому такие алгоритмы могут быть непрактичными из-за чрезмерного времени вычислений.

Методы выбора представлений были классифицированы на основе того, какие алгоритмы они используют для решения проблемы выбора представления, какие ограничения ресурсов они учитывают в процессе выбора представления и какие рамки они используют для получения видов-кандидатов. На основе этой классификации мы рассмотрим большинство методов выбора представления, которые были предложены в литературе:

1. Методы, основанные на детерминированных алгоритмах: в некоторых работах представлен исчерпывающий подход к поиску наилучшего набора представлений, подлежащих реализации. Тем не менее, исчерпывающий поиск не может рассчитать оптимальное решение в разумные сроки. Авторы представляют и анализируют алгоритмы выбора представления в случае старомодных запросов. Они предоставляют алгоритм с интенсивным использованием полиномиального времени для выбора набора представлений

для материализации, который сводит к минимуму затраты на обработку запросов, ограниченного пространством. Однако этот подход не учитывает стоимость обслуживания представления.

2. Методы, основанные на рандомизированных алгоритмах: типичные рандомизированные алгоритмы являются генетическими или используют имитацию отжига. Генетические алгоритмы генерируют решения, используя методы, основанные на естественном эволюционном процессе, такие как отбор, мутация и скрещивание. Стратегия поиска этих алгоритмов очень похожа на биологическую эволюцию. Генетические алгоритмы начинаются со случайной исходной популяции и генерируют новые популяции путем случайно-

го скрещивания и мутации. Самый подходящий человек, которого можно найти, — это решение. Алгоритмы прекращаются, как только в течение определенного периода не происходит дальнейшего улучшения.

3. Методы, основанные на гибридных алгоритмах: гибридные алгоритмы объединяют стратегии детерминированных и рандомизированных алгоритмов в своем поиске, чтобы обеспечить лучшую производительность с точки зрения качества решения. Решения, полученные с помощью детерминированных алгоритмов, используются в качестве начальной конфигурации для алгоритмов имитации отжига или в качестве начальной совокупности для генетических алгоритмов.

ЛИТЕРАТУРА

1. Mukkamala, R., Purna Chandra Rao, V. (2020). Approaches for Efficient Query Optimization Using Semantic Web Technologies. In: Saini, H., Sayal, R., Buyya, R., Aliser, G. (eds) Innovations in Computer Science and Engineering. Lecture Notes in Networks and Systems, vol 103. Springer, Singapore. https://doi.org/10.1007/978-981-15-2043-3_47.
2. Олейникова С.А. Моделирование: учебное пособие / С.А. Олейникова; ФГБОУ «Воронежский государственный технический университет». — Воронеж: Изд-во ВГТУ, 2020. — 127 с.
3. Домбровская Г., Новиков Б., Беликова А. Оптимизация запросов в PostgreSQL / пер. с англ. Д.А. Беликова. — М.: ДМК Пресс, 2022. — 278 с.: ил. ISBN 978-5-97060-963-7.
4. Юе Цюй. Московские годонимы: способы представления лингвокультурной информации в топонимической базе данных обучающего типа // Полилингвистика и транскультурные практики. 2021. Т. 18. № 3. С. 301–309. DOI 10.22363/2618-897X-2021-18-3-301-309.
5. Жалолов О.И., Хаятов Х.У. Понятие SQL и реляционной базы данных // Universum: Технические науки: электрон. научн. журн. 2020. № 6(75). URL: <http://7universum.com/ru/tech/archive/item/9711>.
6. Karras, A., Karras, C., Pervanas, A., Sioutas, S., Zaroliagis, C. (2023). SQL Query Optimization in Distributed NoSQL Databases for Cloud-Based Applications. In: Foschini, L., Kontogiannis, S. (Eds) Algorithmic Aspects of Cloud Computing. ALGOCLOUD 2022. Lecture Notes in Computer Science, vol 13799. Springer, Cham. https://doi.org/10.1007/978-3-031-33437-5_2.
7. An ensemble approach to meta-heuristic algorithms: Comparative analysis And its applications Singh P, Kottath R. (2021) Computers and Industrial Engineering, 162, art. No. 107739 DOI: 10.1016/j.cie.2021.107739.
8. Azgomi, H., Sohrabi, M.K. A novel coral reefs optimization algorithm for materialized view selection in data warehouse environments. Appl Intell 49, 3965–3989 (2019). <https://doi.org/10.1007/s10489-019-01481-w>.
9. Зыкин В.С., Цымблер М.Л. Обновление многотабличных представлений на основе коммутативных преобразований базы данных // Вестник ЮУрГУ. Серия: Вычислительная математика и информатика. 2019. Т. 8, № 2. С. 92–106. DOI: 10.14529/cmse190206.
10. Мамедли Р.Э. Системы управления базами данных: Учебное пособие. Нижневартовск: Издво Нижневартовского государственного университета, 2021. — 214 с.

© Конде Абдул Карим (karimconde112@gmail.com)

Журнал «Современная наука: актуальные проблемы теории и практики»